Clinical Chemistry 65:9
1141–1152 (2019)

Islet Autoantibody Standardization Program 2018 Workshop: Interlaboratory Comparison of Glutamic Acid Decarboxylase Autoantibody Assay Performance

Vito Lampasona,[1*†] David L. Pittman,[2†] Alistair J. Williams,[3] Peter Achenbach,[4] Michael Schlosser,[5,6] Beena Akolkar,[7] William E. Winter,[2] and Participating Laboratories

BACKGROUND: The Islet Autoantibody Standardization Program (IASP) aims to improve the performance of immunoassays measuring type 1 diabetes (T1D)-associated autoantibodies and the concordance of results among laboratories. IASP organizes international interlaboratory assay comparison studies in which blinded serum samples are distributed to participating laboratories, followed by centralized collection and analysis of results, providing participants with an unbiased comparative assessment. In this report, we describe the results of glutamic acid decarboxylase autoantibody (GADA) assays presented in the IASP 2018 workshop.

METHODS: In May 2018, IASP distributed to participants uniquely coded sera from 43 new-onset T1D patients, 7 multiple autoantibody-positive nondiabetic individuals, and 90 blood donors. Results were analyzed for the following metrics: sensitivity, specificity, accuracy, area under the ROC curve (ROC-AUC), partial ROC-AUC at 95% specificity (pAUC95), and concordance of qualitative and quantitative results.

RESULTS: Thirty-seven laboratories submitted results from a total of 48 different GADA assays adopting 9 different formats. The median ROC-AUC and pAUC95 of all assays were 0.87 [interquartile range (IQR), 0.83–0.89] and 0.036 (IQR, 0.032–0.039), respectively. Large differences in pAUC95 (range, 0.001–0.0411) were observed across assays. Of formats widely adopted, bridge ELISAs showed the best median pAUC95 (0.039; range, 0.036–0.041).

CONCLUSIONS: Several novel assay formats submitted to this study showed heterogeneous performance. In 2018, the majority of the best performing GADA immunoassays consisted of novel or established nonradioactive tests that proved on a par or superior to the radiobinding assay, the previous gold standard assay format for GADA measurement.

The Islet Autoantibody Standardization Program (IASP)[8] is a collaborative effort aimed at improving the performance of assays measuring type 1 diabetes (T1D)-associated autoantibodies and the concordance of results between laboratories *(1)*.

IASP is supported by the Immunology of Diabetes Society (IDS) and the NIH, coordinated by an IDS-nominated committee, and run by the University of Florida Pathology Laboratories, Endocrine Autoantibody Laboratory.

IASP organizes international interlaboratory comparison studies in which blinded T1D and control serum samples are tested for T1D-associated autoantibodies by participating laboratories. Centralized collection and analysis of results by the IASP committee provide participants with an unbiased comparison of assay performance. Moreover, IASP fosters the contin-

[1] San Raffaele Diabetes Research Institute, IRCCS San Raffaele Hospital, Milan, Italy; [2] Department of Pathology, University of Florida, Gainesville, FL; [3] Diabetes and Metabolism, Translational Health Sciences, University of Bristol, Bristol, UK; [4] Institute of Diabetes Research, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany; [5] Department of Surgery, University Medical Center Greifswald, Greifswald, Germany; [6] Institute of Pathophysiology, Research Group of Predictive Diagnostics, University of Greifswald, Karlsburg, Germany; [7] Division of Diabetes, Endocrinology, and Metabolic, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD.

* Address correspondence to this author at: San Raffaele Diabetes Research Institute, via Olgettina 60, 20132 Milan, Italy. Fax +39-02-2643-4351; e-mail lampasona.vito@hsr.it.

[†] V. Lampasona and D.L. Pittman contributed equally to this work.

Received March 26, 2019; accepted June 27, 2019.

Previously published online at DOI: 10.1373/clinchem.2019.304196

[8] Nonstandard abbreviations: IASP, Islet Autoantibody Standardization Program; T1D, type 1 diabetes; IDS, Immunology of Diabetes Society; GADA, autoantibodies to glutamic acid decarboxylase GAD65 protein; APPA, average pairwise percent agreement; AC1, Gwet's coefficient of interrater agreement reliability; ROC-AUC, area under the ROC curve; pAUC95, partial ROC-AUC at 95% specificity; OCCC, overall concordance correlation coefficient; RBA, radiobinding assay; LIPS, luciferase immunoprecipitation system; ECL, electrochemiluminescence; CLIA, chemiluminescence immunoassay; LBI, Luminex bead immunoassay; MPNIRF, multiplex plasmonic near-infrared fluorescence; MFRET, multiplex fluorescence energy transfer; ADAP, antibody-dependent agglutination PCR; NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases; DK, digestive and kidney unit; IQR, interquartile range.

uous improvement of T1D autoantibody immunoassays through the dissemination of empirically tested best practice protocols, state-of-the-art reagents, and serum standards.

In this report, we analyze the results of assays for antibodies to glutamic acid decarboxylase 65 (GADA) *(2)* submitted in 2018 to the IASP interlaboratory comparison study and presented at the IASP 2018 workshop held at the 16th Immunology of Diabetes Society Congress in London, UK.

GADAs are found in several neurological and endocrine autoimmune diseases *(3–5)*. In the setting of autoimmune diabetes, GADAs are the most prevalent autoantibody at onset of T1D and the hallmark of latent autoimmune diabetes in adults *(6)*, a slowly progressing form of pancreatic endocrine autoimmunity affecting up to 5% of type 2 diabetes patients. Moreover, GADA measurement is a cornerstone of screening strategies for T1D *(7)*.

The most recent IASP GADA interlaboratory comparison and standardization study took place in 2018, with 37 laboratories from 17 countries in North America, Europe, Asia, and Australia submitting results from 48 different GADA assays, based on 9 different assay formats, after testing blinded samples from 50 cases with T1D or multiple islet autoimmunity and 90 blood donors.

Materials and Methods

STUDY DESIGN

In the 2018 IASP interlaboratory comparison study, participants received sets of the same serum samples consisting of 50 cases (43 sera from new-onset T1D patients and 7 multiple islet autoantibody-positive first-degree relatives of T1D patients enrolled in the TrialNet Ancillary Study—Pathway to Prevention, who during screening showed a transiently altered glucose tolerance test), 90 control samples (all blood donors), and 10 additional samples to be used for substudies unrelated to GADA testing.

The T1D patients had a median age of 14 years (range, 8–47 years) and included 15 females and 28 males, of whom 37 were white, 2 black, 2 of mixed race, and 2 of undisclosed ancestry. The multiple T1D autoantibody-positive participants had a median age of 16 years (range, 12–53 years) and included 4 females and 3 males, all of white ancestry. The blood donors had a median age of 20 years (range, 18–30 years) and included 44 females and 44 males, of whom 69 were white, 19 black, and 2 for whom demographic data were not available.

New-onset T1D samples were contributed by several centers around the world and collected within 14 days of starting insulin treatment. Blood donor samples were collected in the US and included only people without diabetes. All serum samples submitted to the IASP repository were collected upon obtaining written informed consent and with the approval of local ethics committees as required by local regulations according to the ethical principles for medical research involving human participants of the Declaration of Helsinki.

All sera were labeled and distributed as blinded 105-μL frozen aliquots, labeled with an aliquot-specific unique code. Laboratories were free to use any GADA assay format but were asked to provide details of their assay protocol and to report assay results, including raw data, to IASP for analysis using uniform Excel (Microsoft) reporting sheets.

DATA ANALYSIS

We calculated sensitivity and specificity for each assay as the percentage of case sera reported as GADA-positive and as the percentage of blood donor sera reported as negative, respectively.

Adjusted sensitivity 95, i.e., the level of sensitivity corresponding to a specificity of 95%, was derived by placing the threshold for positivity at the 95th percentile of values observed in the blood donor samples in each assay.

Concordance of laboratory-assigned positive or negative scores across assays was expressed as average pairwise percent agreement between assays (APPA), i.e., the average number of times each possible combination of 2 assays agreed on GADA-positive/negative scores divided by the number of samples scored. We tested the occurrence of agreement by pure chance by calculating the first agreement coefficient (AC1) according to Gwet *(8, 9)* and the κ coefficient according to Fleiss *(10)* using the corresponding functions (http://agreestat.com/r_functions.html; Advanced Analytics) in the R language and environment for statistical computing and graphics *(11)*.

Assay performance in discriminating health from disease was analyzed using the area under the ROC curve (ROC-AUC) and the partial ROC-AUC imposing a specificity >95% (pAUC95) *(12)*.

The analysis of interassay antibody titer concordance was performed after ranking of patient and control samples according to autoantibody levels in each assay by calculating the Kendall W rank correlation coefficient *(13)* using the vegan R package (https://CRAN.R-project.org/package=vegan). The significance of differences in mean ranking of selected cases between hybrid solid/liquid phase vs liquid phase-only assays was tested using the Mann–Whitney test. This synthetic index based on ranking was preferred to classical pairwise regression analysis in light of the numerous comparisons to be made and to the presence of systematic differences in measurement of high and low GADA titers across assays, mostly related to the adoption of a variety of alternative algorithms for calculating local arbitrary units.

For laboratories that reported results in WHO units *(14)*, the concordance of antibody titers was evaluated by calculating the overall concordance correlation coefficient (OCCC) according to Barnhart *(15)* using the f.analysis macro *(16)*. The OCCC measures how far the fitted linear relationship of 2 variables x and y deviates from the concordance line (accuracy) and how far each observation deviates from the fitted line (precision) and ranges from 0 to ± 1, where results close to ± 1 stand for near perfect concordance (or perfect discordance) and 0 stands for no correlation.

For all statistical analyses, 2-tailed P values <0.05 were considered as significant.

Results

SUMMARY OF SUBMITTED GADA ASSAY FORMATS

Forty-one laboratories from 17 countries in North America, Europe, Asia, and Australia registered for the IASP2018 interlaboratory comparison study. Of the participating laboratories, 37 submitted results (see Appendix 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol65/issue9) from a total of 48 different GADA assays using 9 different formats.

The submitted assays adopted the following formats: radiobinding assay (RBA) (18 assays, 37.5%) *(17)*; bridge-ELISA (12 assays, 25.0%) *(18)*; luciferase immunoprecipitation system (LIPS) (8 assays, 16.7%) *(19)*; electrochemiluminescence (ECL) (3 assays, 6.3%, of which 1 multiplexed individual GADA measurement of the IgG, IgA, and IgM immunoglobulin classes) *(20)*; chemiluminescence immunoassay (CLIA) (2 assays, 4.2%, of which 1 used a bridge format); Luminex bead immunoassay (LBI) (2 assays, 4.2%); multiplex plasmonic near-infrared fluorescence (MPNIRF) (1 assay, 2.1%) *(21)*; multiplex fluorescence energy transfer (MFRET) (1 assay, 2.1%); and antibody-dependent agglutination PCR (ADAP) (1 assay, 2.1%) *(22)*. Antigen–antibody binding occurred in liquid phase in 29 assays (ADAP, ECL, LIPS, RBA) followed by the capture of immune complexes either through the recovery of immunoglobulins (LIPS, RBA) or tagged antigen (ECL), in hybrid solid/liquid phase in 13 assays (bridge-ELISA and CLIA), solid phase in 3 assays (LBI and MPNIRF), while binding phase was not specified for 1 assay (MFRET). Major characteristics and metrics of each individual assay are reported in Table 1 of the online Data Supplement.

Among RBAs, 5 assays adopted the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) harmonized protocol *(23)* and 4 used truncated GAD65 antigens *(24–26)*. Most RBAs used glutamic acid decarboxylase antigens transcribed and translated in vitro (17 of 18 assays) and radiolabeled with ^{35}S or ^{3}H (in 16 and 1 case, respectively). One assay used ^{125}I iodinated recombinant GAD65 antigen. Results in WHO units *(14)* were reported for all bridge-ELISA assays, 5 RBAs, and 3 LIPS that used digestive and kidney unit (DK) standards *(23)*. Among LIPS, 7 assays used the Nanoluc and 1 used the Renilla luciferase reporters. Most LIPS assays used truncated GAD65 antigens corresponding to glutamic acid decarboxylase amino acids 96 to 585 (GAD96–585, 4 assays) or 188 to 585 (GAD188–585, 2 assays) instead of full-length GAD65 (GAD1–585, 1 assay).

ASSAY SENSITIVITY, SPECIFICITY, AND ACCURACY

The median laboratory-assigned assay sensitivity, specificity, and accuracy were 69% [interquartile range (IQR), 64%–76%], 98.9% (IQR, 96.7%–100%), and 88.6% (IQR, 84.5%–90.7%), respectively (Table 1 and Fig. 1 here and also Table 1 in the online Data Supplement).

ASSAY PERFORMANCE

The median ROC-AUC was 0.87 (IQR, 0.83–0.89) (see Table 1 here and also Table 1 and Fig. 1 in the online Data Supplement). As a more relevant proxy of assay performance based around commonly adopted thresholds for positivity, we calculated also the pAUC95. The median of pAUC95 GADA assays was 0.036 (IQR, 0.032–0.039) against a theoretical maximum of 0.05, and the stratification according to format highlighted a wide heterogeneity of performance across both assays and formats when high specificity was levied (Fig. 2 here and also Fig. 2 in the online Data Supplement).

The median pAUC95 of RBAs was 0.0349 (range, 0.0253–0.0394). The RBAs with pAUC95 above the overall median included assays using a truncated GAD65 antigen corresponding to amino acids 96 to 585 (4 of 4 assays), RBAs using the NIDDK harmonized protocol (2 of 5 assays), and only 2 RBAs using local protocols and full-length GAD65 (2 of 8 assays).

Bridge-ELISA assays' pAUC95 showed a median value of 0.0393 (range, 0.0358–0.0411), always at or above the median of all assays.

ECL assays that, like the bridge-ELISA, are theoretically capable of detecting autoantibodies of any immunoglobulin class showed a pAUC95 of 0.0362 (range, 0.0347–0.0377). A third ECL assay aimed to multiplex the measurement of specific GADA immunoglobulin classes (IgG, IgM, and IgA) and showed variable pAUC95 (ECL-IgG pAUC95 = 0.0304, ECL-IgM pAUC95 = 0.0160, and ECL-IgA pAUC95 = 0.0120).

LIPS assays using a truncated antigen corresponding to GAD65 amino acids 96 to 585 and a nanoluciferase reporter showed a median pAUC95 of 0.0360 (range, 0.0359–0.0368), whereas the LIPS using a Renilla luciferase reporter showed lower performance (pAUC95 = 0.0196). A LIPS assay using full-length GAD65 showed a pAUC95 of 0.034, and 2 LIPS assays using a truncated

Table 1. IASP2018 summary of GADA assay performance.

Format	Number of assays	Sensitivity	Specificity	Accuracy	ROC-AUC	pAUC95
ADAP	1	80	100	92.9	0.96	0.041
Bridge-ELISA	12	78.0 (68.0–80.0)	100.0 (95.5–100.0)	91.4 (88.6–92.8)	0.89 (0.88–0.94)	0.039 (0.036–0.041)
RBA						
Truncated GAD65 aa96–585	4	69.0 (66.0–74.0)	100.0 (95.6–100.0)	88.2 (87.9–89.3)	0.84 (0.82–0.88)	0.037 (0.036–0.039)
Local protocol	9	66.0 (50.0–72.0)	97.8 (95.7–100.0)	87.1 (82.1–88.6)	0.84 (0.75–0.89)	0.034 (0.029–0.037)
Harmonized protocol	5	64.0 (58.0–82.0)	98.9 (94.4–100.0)	85.0 (81.4–92.9)	0.87 (0.77–0.91)	0.032 (0.025–0.039)
LIPS						
NLuc reporter GAD65 aa96–585	4	70.0 (60.0–76.0)	99.4 (98.9–100.0)	88.9 (85.7–90.7)	0.87 (0.86–0.89)	0.036 (0.036–0.037)
NLuc reporter GAD65 aa188–585	2	71.0 (68.0–74.0)	98.3 (97.8–98.9)	88.6 (87.9–89.3)	0.86 (0.86–0.86)	0.035 (0.034–0.035)
NLuc reporter GAD65 aa1–585	1	72.0	98.9	89.3	0.84	0.034
FLuc reporter GAD65 aa96–585	1	54	96.7	81.4	0.79	0.019
ECL						
ECL	2	78.0 (78.0–82.0)	97.2 (95.6–98.9)	91.1 (89.3–92.8)	0.89 (0.88–0.91)	0.036 (0.035–0.038)
ECL IgG specific	1[a]	66.0	96.7	85.7	0.88	0.030
ECL IgM specific	1[a]	30.0	97.8	73.6	0.70	0.016
ECL IgA specific	1[a]	28.0	97.8	72.9	0.67	0.012
CLIA	2	72.0 (68.0–76.0)	96.7 (93.3–100.0)	87.9 (84.3–91.4)	0.88 (0.85–0.92)	0.035 (0.031–0.039)
MPNIRF	1	50.0	95.6	79.3	0.87	0.021
LBI	2	36.0 (36.0–36.0)	87.2 (86.7–87.8)	68.9 (68.6–69.3)	0.58 (0.57–0.59)	0.011 (0.011–0.012)
MFRET	1	2.0	97.8	63.6	0.18	0.001
All assays: median (IQR)	48	69.0 (64.0–76.0)	98.9 (96.7–100.0)	88.6 (84.5–90.7)	0.87 (0.83–0.89)	0.036 (0.032–0.039)

[a] Multiplexed ECL assay.

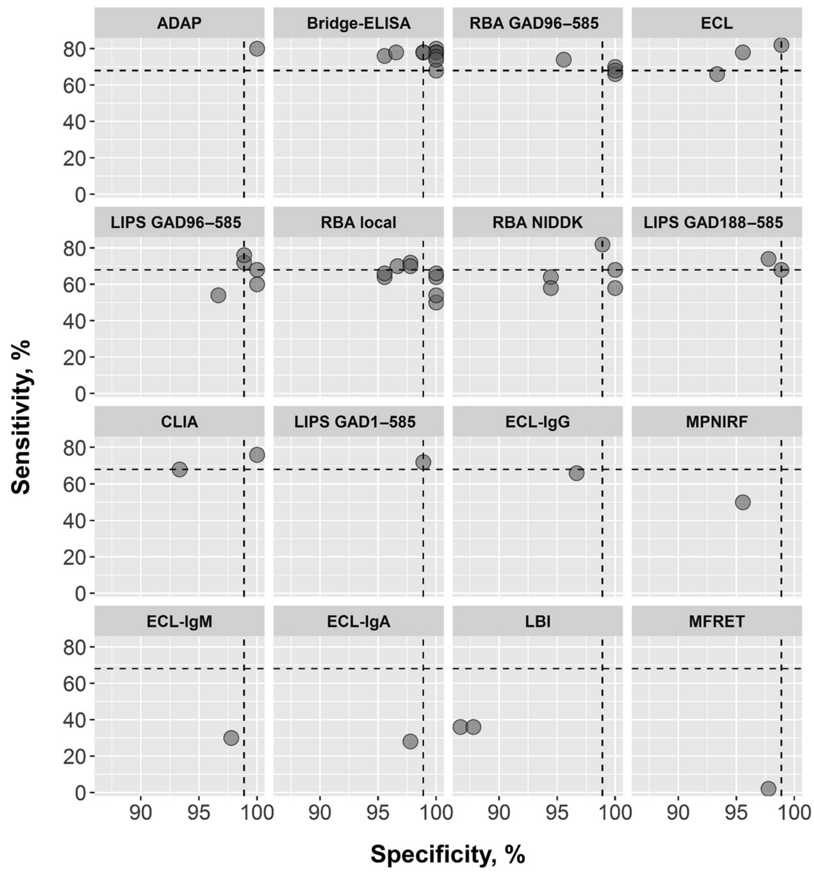

Fig. 1. Scatter plots of sensitivity and specificity of GADA assays based on laboratory-assigned GADA-positive or -negative scores for 50 cases and 90 controls.

Filled circles stand for individual assays. Dashed lines mark the median sensitivity and specificity of all assays. Assays are categorized according to format and its variants. Categories are sorted by their median assay performance.

antigen corresponding to GAD65 amino acids 188 to 585 had a median pAUC95 of 0.034 (range, 0.034 to 0.035).

Among assays that used formats submitted for the first time to the IASP interlaboratory comparison study, the best performance was achieved by the ADAP assay (pAUC95 = 0.0411), followed by 2 assays using the CLIA format with a median pAUC95 of 0.035 (range, 0.031–0.038), whereas a poorer performance was observed for the 3 remaining assays adopting the MPNIRF (pAUC95 = 0.014), LBI (pAUC95 = 0.006), and MFRET (pAUC95 = 0.001) formats, respectively.

CONCORDANCE OF LABORATORY-ASSIGNED POSITIVE/NEGATIVE SCORES

In cases, the APPA of positive/negative scores across all assays was 83.16% and the first-order agreement coefficient AC1 *(9)* was 0.69. Concordance increased when the analysis was limited to assays with pAUC95 above the median (APPA = 90.82%; AC1 = 0.85) or when assays using the same format and/or antigens were compared (APPA range = 84.0%–96.31%; AC1 range = 0.73–0.94) (see Table 2 in the online Data Supplement).

In control samples, the APPA and AC1 coefficient across all assays were 96.10% and 0.96, respectively. Similar to cases, both agreement measures increased when assays with pAUC95 above the median (APPA = 97.52%; AC1 = 0.97) or assays using the same format and/or antigens were compared (APPA range = 93.3%–98.2%; AC1 range = 0.92–0.98) with the exception of ECL and CLIA assays (see Table 3 in the online Data Supplement).

The results of the Fleiss κ concordance coefficient *(10)*, an alternative metric of agreement, were invariably lower (range, −0.03 to 0.49), an observation consistent with the majority of control samples being scored

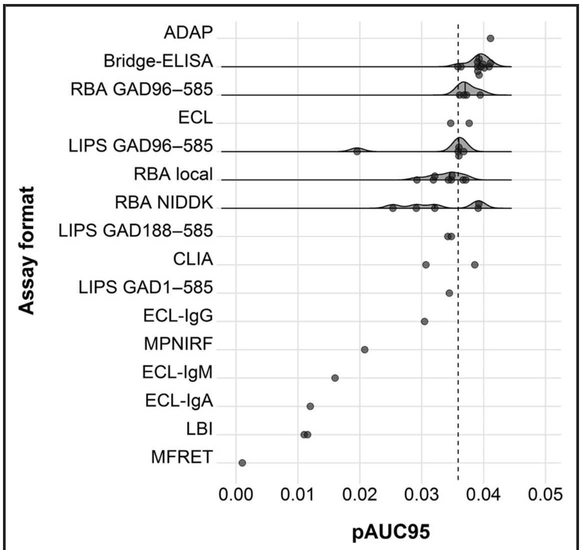

Fig. 2. Performance of GADA assays in IASP2018.
Shown are the pAUC95 of each assay (gray filled dots) and the probability density estimates of the pAUC95 distribution. Assays are grouped by format and its variants and the groups sorted according to their median pAUC95. The dashed line marks the median pAUC95 of all assays.

GADA-positive sporadically in only a small fraction of assays *(27)*.

The analysis of laboratory-assigned positive/negative scores highlighted the existence of format-associated patterns. A subset of cases was recognized as GADA-positive predominantly by ADAP and bridge-ELISA assays (samples IDS324, IDS309, IDS290, IDS312, and IDS337) and to a lower extent by some ECL (samples IDS309, IDS290, IDS312, and IDS337) or LIPS assays (samples IDS312 and IDS337) but not by the majority of RBAs (Fig. 3). Conversely, a subset of controls was recognized as GADA-positive exclusively by a minority of RBAs (samples TS23727, N59416, and N53371) or LIPS assays (sample S8650) (Fig. 4).

To evaluate the impact of local threshold selection criteria on assay sensitivity and specificity, the data were reanalyzed after imposing a 95% specificity onto IASP2018 controls. The newly assigned positive/negative scores highlighted an improved concordance between bridge-ELISAs or ECLs and RBAs or LIPS, with all samples previously recognized as positives exclusively by the hybrid solid/liquid phase bridge-ELISA format scoring positive also in liquid phase assays (see Fig. 3 in the online Data Supplement). In control samples, the use of these novel thresholds made more apparent the presence of weak format and/or antigen-specific reactivities that lead to frequent positive scores for subsets of control samples in local or harmonized protocol RBAs (samples TS23727, N59416, and N53371) and RBAs or LIPS assays using truncated GAD96–585 (samples LQ23340, N59534, and S8650) (see Fig. 4 in the online Data Supplement).

CONCORDANCE OF AUTOANTIBODY TITER RANKS

The interassay concordance of antibody titer was evaluated by first ranking sera in each assay (Fig. 5A here and also Figs. 5 and 6 in the online Data Supplement) followed by calculation of the Kendall W ranking agreement coefficient *(13)*. The W coefficients across all assays were 0.80 and 0.13 in cases and control samples, respectively (see Tables 2 and 3 in the online Data Supplement).

The exclusion from the analysis of assay formats with the 4 lowest pAUC95 led to a modest increase of the agreement coefficient in both cases and controls (W, 0.85 and 0.14, respectively), while limiting the analysis to assays with performance above the overall median pAUC95 showed again an increase of W in cases (W = 0.93) but only a marginal improvement in controls (W = 0.15), suggesting that higher concordance in cases was partially correlated with assay performance.

Concordance of GADA titer ranks increased among assays using the same format both in cases and, at least for assays other than the bridge-ELISA, also in control samples (see Tables 2 and 3 and Figs. 5 and 6 in the online Data Supplement).

A marginal increase of W in cases after excluding assays measuring only IgM or IgA antibodies (W = 0.82) was observed, indicating that discrepancies of GADA titer ranks across assays were not exclusively associated with measurement of immunoglobulins of a class other than IgG.

Stratification of assays according to antibody/antigen binding in hybrid solid/liquid phase, i.e., bridge-ELISAs, or liquid phase, i.e., including ECL, RBA, and LIPS, highlighted a significant difference between hybrid solid/liquid or liquid phase assays in mean GADA titer ranks assigned to a subset of T1D sera (IDS285, IDS301, IDS319, and IDS325; Mann–Whitney test: all $P < 0.0001$).

CONCORDANCE OF AUTOANTIBODY UNITS

The concordance of arbitrary units assigned to cases was assessed across assays expressing results in WHO units. These included bridge-ELISAs, all commercial assays using standards calibrated against the National Institute of Biological Standards and Control 97/550 WHO reference serum *(14)*, and a subset of RBA and LIPS using the NIDDK DK standards, also calibrated against the WHO reference serum. Across all assays, interassay concordance of WHO units was relatively low with an OCCC of 0.4870.

After assay stratification according to antibody/antigen binding in hybrid solid/liquid phase or liquid phase, concordance of antibody titers increased across liquid phase as-

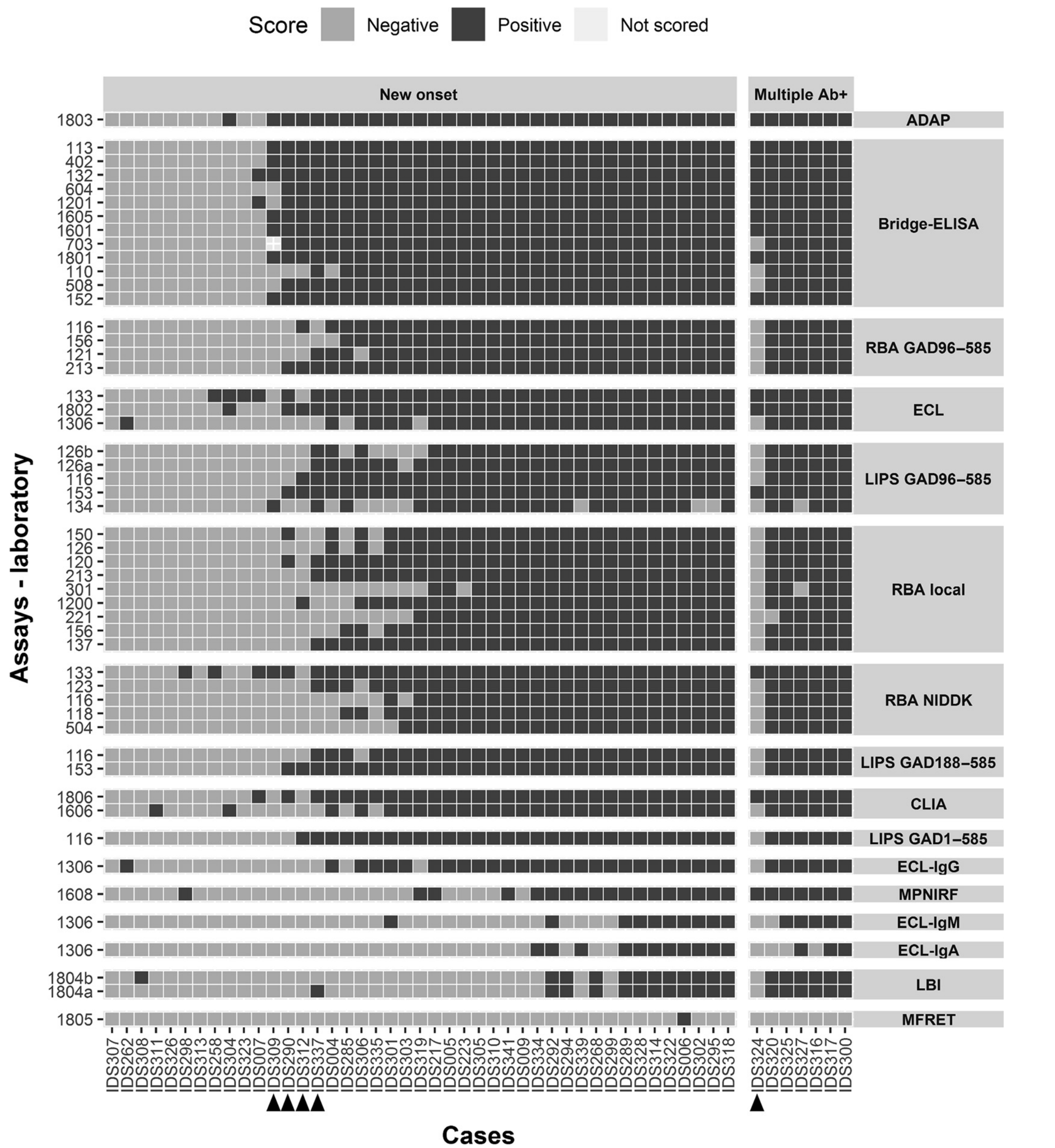

Fig. 3. Tile map of GADA-positive (dark gray) or -negative (light gray) scores assigned by laboratories to cases.
Assays are grouped by format and its variants and the groups sorted according to their median pAUC95. Cases are ordered by median GADA titer rank. A cumulative score of individual Ig class results was added for lab 1306 in the ECL group. Cases with distinctive patterns discussed in the text are indicated by a black triangle.

says adopting the NIDDK calibrators (OCCC = 0.82) compared with bridge-ELISAs (OCCC = 0.45) (Fig. 5B).

Stratification of assays confirmed a significant difference between hybrid solid/liquid or liquid phase assays in the GADA titers assigned to a subset of T1D sera (IDS285, IDS301, IDS319, and IDS325) (Mann–Whitney test: $P < 0.0001$, $P < 0.0011$, $P < 0.0001$, and $P < 0.0008$, respectively) (Fig. 5C).

Discussion

Workshops aimed at the standardization of T1D-associated autoantibodies were originally established by an international community of researchers trying to address early discrepancies in results between laboratories measuring islet cell antibodies *(28, 29)*. The success of this initiative led to the recognition of the

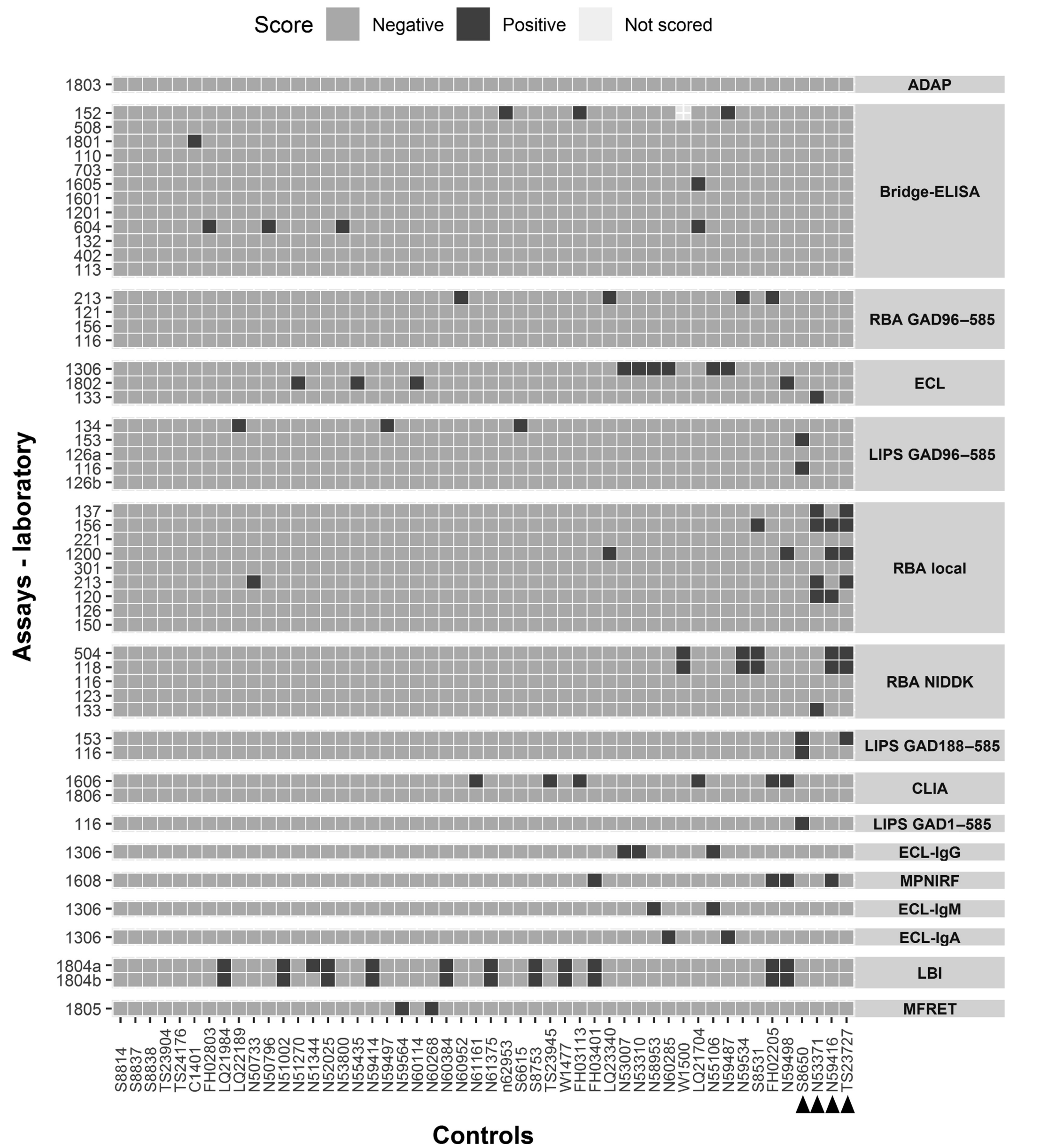

Fig. 4. Tile map of GADA-positive (dark gray) or -negative (light gray) scores assigned by laboratories to controls.
Assays are grouped by format and its variants and the groups sorted according to their median pAUC95. Omitted are 40 controls negative in all assays. A cumulative score of individual Ig class results was added for lab 1306 in the ECL group. Controls with distinctive patterns discussed in the text are indicated by a black triangle.

crucial role of assay format and protocols in determining the results of T1D immunoassays *(30)* and to the implementation of a harmonization and assessment/validation program for all major T1D autoantibodies.

Following the molecular identification of GAD65 as a major T1D autoantigen, harmonization workshops of GADA measurement accompanied the continuous development, testing, and validation of a variety of GADA immunoassays *(31)*. Early interlaboratory comparison studies led to the emergence of the liquid phase immunoprecipitation RBA as a widely implemented de facto gold standard for GADA measurement *(32)*.

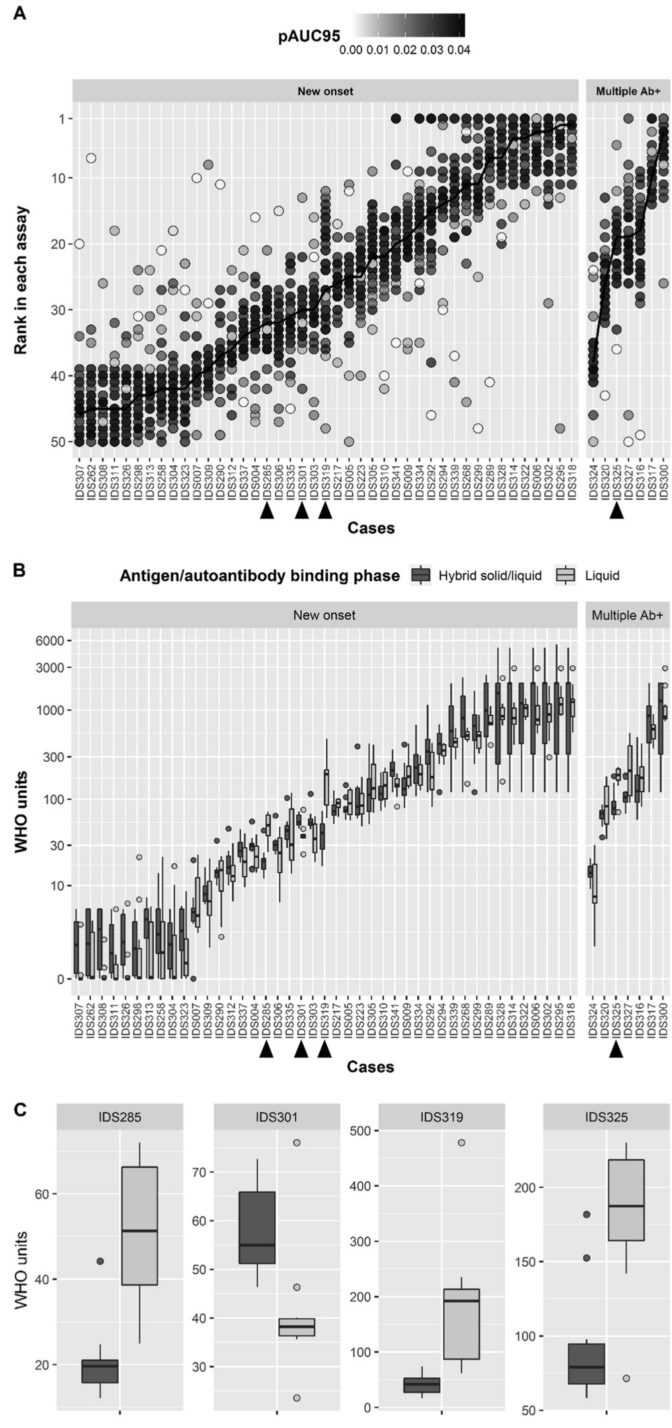

Fig. 5. Analysis of GADA titers' concordance in IASP2018.

(A), GADA titer ranks of each case in different assays (circles, filling is proportional to pAUC95). The black line shows the median rank of all assays. (B), Boxplots of GADA WHO units in each case. Assays are grouped by binding phase (dark gray: hybrid solid/liquid all bridge-ELISA assays; light gray: liquid, include ECL, LIPS, and RBA assays). (C), Enlargement showing samples with significant differences in WHO units between assay groups. Cases selected in (C) are indicated in (A) and (B) by a black triangle.

RBAs not only repeatedly demonstrated good sensitivity and specificity in several interlaboratory comparisons *(33)* but also allowed laboratories to produce radiolabeled GAD65 using a simple in vitro system and, because in RBA antigen/antibody binding occurs in liquid phase, assuaged the concern for the preservation of GADA conformational epitopes without the disruption often associated with antigen adsorption to a solid phase *(34)*.

The IASP2018 interlaboratory comparison study saw the continued implementation of the RBA format by the majority of laboratories, together with an expanded adoption of alternative nonradioactive formats, allowing for a comprehensive comparison between classical RBA and other assay formats in terms of diagnostic, rather than analytical, sensitivity, specificity, accuracy, concordance, and performance.

Regarding the analysis of GADA assay performance in IASP2018, in addition to calculating the ROC-AUC curve, possibly the most widely used metric of a diagnostic test performance, we evaluated the pAUC95. This approach improves one of the major limitations and confounders of using the total ROC-AUC, i.e., the inclusion in the analysis of regions of low assay specificity that are not clinically relevant *(12)*.

The characteristics of RBAs submitted to IASP2018 showed the presence of several alternative protocols differing with regard to assay buffers, washing method (i.e., centrifugation vs filtration), amount of test serum and antigen, radiolabel (^{35}S, ^{125}I, ^{3}H), and the use of full-length or truncated GAD65 antigens *(24–26)*. Unsurprisingly, RBAs presented a large degree of variability in performance, with only a few RBAs using protocols developed in-house achieving a pAUC95 above the median of all assays.

Among these protocol variables, the use of truncated GAD65 antigen corresponding to amino acids 96 to 585 of the full-length protein was associated with better pAUC95 compared with most RBAs using full-length GAD65 in both RBAs and LIPS assays, another immunoprecipitation-based format.

In IASP2018, the second most widely adopted immunoassays were commercial bridge-ELISAs. Bridge-ELISAs showed the most homogeneous and highest pAUC95 of all assays, save for the novel PCR-based ADAP format, and proved more sensitive and specific overall than most other assays, with the few positive scores discrepant with those of high-performance RBAs and LIPS assays, essentially limited to low titer GADA samples.

We then reassessed positive scores after applying a threshold based on a common predefined specificity of 95% to all assays, a threshold that, although not ideal if applied to population screening, facilitated comparisons across this limited sample set. This reanalysis showed that discrepancies in low titer GADA cases were resolved in the majority of LIPS assays and RBAs, suggesting that most liquid phase assays in IASP2018 generally adopted more conservative thresholds compared with the bridge-ELISA.

The consistent detection by some RBAs and LIPS assays of slightly increased antibody binding in some specific blood donor sera from the IASP2018 sample set suggests that threshold selection in these assays might have been driven by the presence of low, non–disease-specific antigen binding in some local control samples. Based on previous publications, it can be speculated that at least part of this binding might be attributed to low affinity antibodies *(25, 35)*.

The comparison of quantitative GADA results in IASP2018 was complicated by the implementation of a variety of local nonstandardized arbitrary units and calculation algorithms to express results in place of the international units endorsed by the WHO. Previous workshops addressed this source of interassay variability by distributing the WHO standard serum and encouraging reporting of GADA in WHO units. Because the WHO reference consists of a strongly autoantibody-positive human serum and constitutes an intrinsically finite resource, its distribution was meant to be used for recalibration of local standards followed by conversion of local to WHO units. However, after encouraging preliminary results *(36)*, further analyses did not confirm the same level of concordance and reproducibility of WHO units across laboratories *(37)*. This prompted the design of a common protocol, including an alternative set of GADA serum standards calibrated against the WHO serum with the aim of harmonizing RBAs used by NIDDK-sponsored consortia *(23)*.

In IASP2018 the majority of assays reporting results in WHO units consisted of bridge-ELISAs, whereas only 2 RBAs using a protocol developed in-house did so. NIDDK standards were used by only 3 laboratories, and these submitted results for RBAs using the NIDDK harmonized protocol and/or LIPS assays.

Overall, although IASP2018 GADA assays showed a reasonable interassay agreement of GADA titer ranks among the assays using WHO units, at least in cases if not in control individuals, we observed a relatively lower concordance of attributed GADA titers, particularly in a subset of cases, despite the use of standardized protocols and centrally prepared calibrators.

Moreover, a dichotomy in GADA titers attributed to selected sera was evident when assays were grouped according to antigen/antibody binding phase into hybrid solid/liquid or just liquid phase assays. Although the underlying reason for this behavior remains to be clarified, multiple potential causes can be hypothesized, like epitope alterations following the addition of tags (e.g., biotin residues in the bridge-ELISA or luciferase enzyme in LIPS), subtle differences in primary amino acid sequence, posttranslational modifications in different expression systems, and the adsorption of antigens onto solid surfaces in the bridge-ELISA. The frequency of this phenomenon remains to be ascertained along with its potential impact, if any, on autoantibody-based screening strategies, which are currently mostly based on RBAs.

Several novel immunoassays were submitted to IASP2018, a development likely spurred by the continuous legislative and logistic pressure against the use of radioactive substances and the expected future implementation of antibody-based population screening programs for T1D. Among these assays, the PCR-based ADAP assay achieved both high sensitivity and perfect specificity, whereas the rest showed variable performance that in some cases was dramatically inferior to that of more mature formats.

In conclusion, the IASP2018 results depict the field of GADA measurement as both mature, with numerous assays achieving high performance, but also in relative flux, with the active development and deployment of novel immunoassays dispensing altogether with the need for radio isotopic tracers. We believe that these results confirm the usefulness of harmonization programs, not only as providers of unbiased diagnostic performance assessment to participants but also as an arena in which research laboratories and companies can learn valuable lessons for improving immunoassays for T1D autoantibodies.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

W.E. Winter, administrative support, provision of study material or patients.

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: Support received by IASP study through a TEDDY *(38)* grant administered at the University of South Florida (Tampa, FL) and by the Type 1 Diabetes TrialNet/IASP Ancillary studies arm *(39)*.
Expert Testimony: None declared.
Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

Acknowledgments: The authors thank all the people and the clinicians who contributed blood samples without whom generosity and effort the IASP workshops and collaborative studies could not take place.

Participating Laboratories and Contacts for the IASP 2018 GADA interlaboratory comparison study:

K. Watson, Royal Melbourne Hospital, Victoria, Australia;
I. Weets, Uzbrussel Vub, Clinical Biology of Diabetes-Diabetes Research Center, Brussel, Belgium;
Y. Tao, Department of Endocrinology, First Affiliated Hospital of Nanjing Medical University, Nanjing, China;
V. Chen, Laboratory of Snibe, Shenzhen, China;
Y. Yang, Shenzhen YHLO Biotech Co., Ltd, Shenzhen, China;
R. Uibo, K. Reimand, University of Tartu, Department of Immunology, Tartu, Estonia;
M. Knip, T. Härkönen, Children's Hospital, Scientific Laboratory, University of Helsinki, Helsinki, Finland;
L. Chatenoud, Laboratoire D'immunologie Biologique-Hôpital Necker-Enfants Malades Paris, Paris, France;
P. Achenbach, Institute of Diabetes Research, Helmholtz Zentrum München, Neuherberg, Germany;
S. Neidhoefer, AESKU.Diagnostics Gmbh & Co.KG, Wendelsheim, Germany;
M. Schlosser, Department of Surgery, University Medical Center Greifswald, Karlsburg, Germany;
V. Lampasona, San Raffaele Diabetes Research Institute, IRCCS Istituto Scientifico San Raffaele, Milano, Italy;
E. Kawasaki, Diabetes Center, Shin-Koga Hospital, Kurume, Japan;
M. R. Batstra, T. Cieremans, Reinier De Graaf Groep, Dept of Medical Immunology, Delft, Netherlands;
B. Almås, The Hormone Laboratory, Haukeland University Hospital, Bergen, Norway;
K. S. Opsion, Hormone Laboratory, Oslo University Hospital, Oslo, Norway;
K. Wyka, Immunopathology And Genetics, Medical University, Lódz, Poland;
L. Castaño, Diabetes Research Laboratory, Biocruces-Hospital Universitario Cruces, Barakaldo, Spain;
A. Ramelius, Diabetes And Celiac Disease Research Unit, Lund University, Malmö, Sweden;
I. Johansson, Clinical & Experimental Research, Division of Pediatrics, Linköping, Sweden;
A. Williams, Diabetes & Metabolism, Learning & Research, University of Bristol, Bristol, UK;
J. Furmaniak, FIRS Laboratories, RSR Ltd, Cardiff, UK;
T. McDonald, Royal Devon And Exeter NHS Foundation Trust, Exeter, UK;
K. McLaughlin, OCDEM, University of Oxford, Oxford, UK;
M. Christie, University of Lincoln, UK;
A. Metz, Wave 80 Biosciences, Inc., Gaithersburg, MD, USA;
A. Mathew, Meso Scale Diagnostics, LLC., Gaithersburg, MD, USA;
C. Hampe, University of Washington, Seattle, WA, USA;
C. Lu, Meso Scale Discovery, Gaithersburg, MD, USA;
C. Wasserfall, University of Florida, Gainesville, FL, USA;
C. Mann, Quest Diagnostics Nichols Institute, San Juan Capistrano, CA, USA;
D. Pittman, University of Florida Health Pathology Laboratories, University of Florida Health Pathology Laboratories, Gainesville, FL;
J. S. Ananta, Nirmidas Biotech INC, Palo Alto, CA, USA;
L. Yu, Barbara Davis Center, Aurora, CO, USA;
M. Mamula, L2 Diagnostics at Yale University, New Haven, CT, USA;
P. Robinson, Enable Biosciences, INC, San Francisco, CA, USA;
V. Gaur, Northwest Lipid Metabolism and Diabetes Research Laboratories, Seattle, WA, USA;
W. A. Hagopian, Pacific Northwest Diabetes Research Institute, Seattle, WA, USA.

References

1. Bingley PJ, Williams AJK. Validation of autoantibody assays in type 1 diabetes: workshop programme. Autoimmunity 2004;37:257–60.
2. Baekkeskov S, Aanstoot HJ, Christgau S, Reetz A, Solimena M, Cascalho M, et al. Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. Nature 1990;347:151–6.
3. Solimena M, Folli F, Denis-Donini S, Comi GC, Pozza G, De Camilli P, et al. Autoantibodies to glutamic acid decarboxylase in a patient with stiff-man syndrome, epilepsy, and type I diabetes mellitus. N Engl J Med 1988; 318:1012–20.
4. Saiz A, Blanco Y, Sabater L, González F, Bataller L, Casamitjana R, et al. Spectrum of neurological syndromes associated with glutamic acid decarboxylase antibodies: diagnostic clues for this association. Brain J Neurol 2008; 131:2553–63.
5. Soderbergh A, Myhre AG, Ekwall O, Gebre-Medhin G, Hedstrand H, Landgren E, et al. Prevalence and clinical associations of 10 defined autoantibodies in autoimmune polyendocrine syndrome type I. J Clin Endocrinol Metab 2004;89:557–62.
6. Groop LC, Zimmet P, Rowley MJ, Knowles W, Mackay IR. Antibodies to glutamic acid decarboxylase reveal latent autoimmune mellitus in adults with a non-insulin-dependent onset of disease. Diabetes 1993;42:4.
7. Krischer JP, Lynch KF, Lernmark Å, Hagopian WA, Rewers MJ, She J-X, et al. Genetic and environmental interactions modify the risk of diabetes-related autoimmunity by 6 years of age: the TEDDY study. Diabetes Care 2017;40:1194–202.
8. Gwet KL. Testing the difference of correlated agreement coefficients for statistical significance. Educ Psychol Meas 2016;76:609–37.
9. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013;13:61.
10. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378–82.
11. R Development Core Team. R: a language and environment for statistical computing. http://www.r-project.org (Accessed July 2019).
12. Ma H, Bandos AI, Gur D. On the use of partial area under the ROC curve for comparison of two diagnostic tests. Biom J 2015;57:304–20.
13. Kendall MG, Babington Smith B. The problem of *m* rankings. Ann Math Stat 1939;10:275–87.
14. Mire-Sluis AR, Gaines Das R, Lernmark A. The World Health Organization International Collaborative Study for islet cell antibodies. Diabetologia 2000;43: 1282–92.
15. Barnhart HX, Haber M, Song J. Overall concordance correlation coefficient for evaluating agreement among multiple observers. Biometrics 2002;58:1020–7.
16. Crawford SB, Kosinski AS, Lin H-M, Williamson JM, Barnhart HX. Computer programs for the concordance correlation coefficient. Comput Methods Programs Biomed 2007;88:62–74.
17. Petersen JS, Hejnaes KR, Moody A, Karlsen AE, Marshall MO, Høier-Madsen M, et al. Detection of GAD65 antibodies in diabetes and other autoimmune diseases using a simple radioligand assay. Diabetes 1994;43: 459–67.
18. Brooking H, Ananieva-Jordanova R, Arnold C, Amoroso M, Powell M, Betterle C, et al. A sensitive non-isotopic assay for GAD65 autoantibodies. Clin Chim Acta 2003; 331:55–9.
19. Burbelo PD, Groot S, Dalakas MC, Iadarola MJ. High definition profiling of autoantibodies to glutamic acid decarboxylases GAD65/GAD67 in stiff-person syndrome. Biochem Biophys Res Commun 2008;366:1–7.
20. Miao D, Guyer KM, Dong F, Jiang L, Steck AK, Rewers M, et al. GAD65 autoantibodies detected by electrochemiluminescence assay identify high risk for type 1 diabetes. Diabetes 2013;62:4174–8.
21. Zhang B, Kumar RB, Dai H, Feldman BJ. A plasmonic chip for biomarker discovery and diagnosis of type 1 diabetes. Nat Med 2014;20:948–53.
22. Tsai C, Robinson PV, Spencer CA, Bertozzi CR. Ultrasensitive antibody detection by agglutination-PCR (ADAP). ACS Cent Sci 2016;2:139–47.
23. Bonifacio E, Yu L, Williams AK, Eisenbarth GS, Bingley PJ, Marcovina SM, et al. Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for National Institute of Diabetes and Digestive and Kidney Diseases consortia. J Clin Endocrinol Metab 2010;95: 3360–7.
24. Williams AJK, Lampasona V, Schlosser M, Mueller PW, Pittman DL, Winter WE, et al. Detection of antibodies directed to the N-terminal region of GAD is dependent on assay format and contributes to differences in the specificity of GAD autoantibody assays for type 1 diabetes. Diabetes 2015;64:3239–46.
25. Williams AJK, Lampasona V, Wyatt R, Brigatti C, Gillespie KM, Bingley PJ, et al. Reactivity to N-terminally truncated GAD_{65}(96–585) identifies GAD autoantibodies that are more closely associated with diabetes progression in relatives of patients with type 1 diabetes. Diabetes 2015;64:3247–52.
26. Achenbach P, Hawa MI, Krause S, Lampasona V, Jerram ST, Williams AJK, et al. Autoantibodies to N-terminally truncated GAD improve clinical phenotyping of individuals with adult-onset diabetes: action LADA 12. Diabetologia 2018;61:1644–9.
27. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. resolving the paradoxes. J Clin Epidemiol 1990;43:551–8.
28. Gleichmann H, Bottazzo GF. Progress toward standardization of cytoplasmic islet cell-antibody assay. Diabetes 1987;36:578–84.
29. Boitard C, Bonifacio E, Bottazzo GF, Gleichmann H, Molenaar J. Immunology and Diabetes Workshop: report on the Third International (Stage 3) Workshop on the Standardisation of Cytoplasmic Islet Cell Antibodies. Held in New York, New York, October 1987. Diabetologia 1988;31:451–2.
30. Liu E, Eisenbarth GS. Accepting clocks that tell time poorly: fluid-phase versus standard ELISA autoantibody assays. Clin Immunol 2007;125:120–6.
31. Schmidli RS, Colman PG, Bonifacio E, Bottazzo GF, Harrison LC. High level of concordance between assays for glutamic acid decarboxylase antibodies. Diabetes 1994;43:5.
32. Verge CF, Stenger D, Bonifacio E, Colman PG, Pilcher C, Bingley PJ, et al. Combined use of autoantibodies (IA-2 autoantibody, GAD autoantibody, insulin autoantibody, cytoplasmic islet cell antibodies) in type 1 diabetes: Combinatorial Islet Autoantibody Workshop. Diabetes 1998;47:1857–66.
33. Törn C, Mueller PW, Schlosser M, Bonifacio E, Bingley PJ, participating laboratories. Diabetes Antibody Standardization Program: evaluation of assays for autoantibodies to glutamic acid decarboxylase and islet antigen-2. Diabetologia 2008;51:846–52.
34. Schwab C, Bosshard HR. Caveats for the use of surface-adsorbed protein antigen to test the specificity of antibodies. J Immunol Methods 1992;147:125–34.
35. Miao D, Steck AK, Zhang L, Guyer KM, Jiang L, Armstrong T, et al. Electrochemiluminescence assays for insulin and glutamic acid decarboxylase autoantibodies improve prediction of type 1 diabetes risk. Diabetes Technol Ther 2015;17:119–27.
36. Bingley PJ, Bonifacio E, Mueller PW. Diabetes antibody standardization program: first assay proficiency evaluation. Diabetes 2003;52:1128–36.
37. Bingley PJ, Williams AJ, Colman PG, Gellert SA, Eisenbarth G, Yu L, et al. Measurement of islet cell antibodies in the Type 1 Diabetes Genetics Consortium: efforts to harmonize procedures among the laboratories. Clin Trials 2010;7:S56–64.
38. TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. Ann N Y Acad Sci 2008;1150:1–13.
39. Skyler JS, Greenbaum CJ, Lachin JM, Leschek E, Rafkin-Mervis L, Savage P, et al. Type 1 Diabetes TrialNet–an international collaborative clinical trials network. Ann N Y Acad Sci 2008;1150:14–24.